THE NUMBER OF OBSERVATIONS UPON WHICH A LIMEN MAY BE BASED

By E. G. Boring, Cornell University

Dr. Fernberger has recently discussed the "least number of determinations upon which the measurement of an individual's sensitivity can safely be based."[1] He concludes that with lifted weights "50 determinations is the smallest number upon which an accurate judgment can be based," and argues that the anthropometrist, who seeks to combine accuracy of measurement with economy of time, cannot afford to take less than 50 series. This limit is fixed by the curve of practice; the decrease in the interval of uncertainty when 25 series are increased to 50 is great; the further decrease in the interval when 100 series are taken is comparatively small.

The minimal number of cases upon which a limen can be reliably based must, however, depend upon something more than practice. Its determination must take into account the number of cases which will give statistical significance to the final average or limen. If this were not the case, Fernberger would do well to base his limen upon the second 25 series rather than upon the entire 50, and still better upon the last 10 of the 50, for practice would be maximal in the final series.

It is not possible, moreover, to place an arbitrary limit upon the number of observations; for the required number varies with the use that is to be made of the limen. A limen has scientific or anthropometric value only as it may be compared with other limens. A comparison, in terms of some such measure, of the savage with the American, of the right arm with the left, or of the morning with the afternoon, has in it the rudiments of a scientific conclusion. We are interested, therefore, not in limens, but in *differences between* limens. We want to be sure that the savage is more sensitive than the civilized man, or the right arm than the left. Even when the limen is determined for a practical purpose its com-

[1] S. W. Fernberger, The Effects of Practice in Its Initial Stages in Lifted Weight Experiments and Its Bearing upon Anthropometric Measurements, this Journal, 27, April, 1916, 261-272. See especially 270 ff.

parison with some practical norm is still implied. And further we may lay down the general rule that the greater the difference we are seeking to establish, the fewer are the observations necessary to establish it. A dozen observations will demonstrate a difference in the limen of duality between the back and the finger-tip; five hundred may be necessary to show a difference between the two forearms.

Limens of dual discrimination (mm.) for the forearm (longitudinal) and the lower eyelid; differences between these limens; and probable correctness of the differences, arranged for various groupings of the 100 series taken in each case.

Series	Forearm			Eyelid			Difference		
	L	h	$P.E._L$	L	h	$P.E._L$	D	$\frac{D}{P.E._D}$	Probable correctness
1st ten........	7.55	.040	1.96	5.70	.152	.456	1.85	0.90	.7270
2nd ten........	12.47	.083	1.15	5.76	.212	.403	6.71	5.52	.9999
3rd ten........	14.08	.046	1.60	5.29	.134	.473	8.79	5.26	.9998
4th ten........	7.61	.054	1.40	5.29	.166	.394	2.32	1.60	.8597
5th ten........	12.16	.087	1.30	6.81	.397	.255	5.35	4.05	.9968
6th ten........	14.00	.077	1.02	5.67	.207	.394	8.33	7.64	1.0000
7th ten........	12.74	.039	2.20	5.13	.324	.260	7.61	3.59	.9922
8th ten........	12.02	.113	1.11	5.59	.113	.580	6.43	5.14	.9997
9th ten........	9.66	.084	1.03	6.17	.248	.351	3.49	3.20	.9845
10th ten........	12.51	.110	0.84	5.49	.125	.570	6.80	7.68	1.0000
1st ten........	7.55	.083	1.15	5.70	.152	.403	1.85	0.90	.7269
1st twenty....	10.94	.061	0.91	5.58	.225	.225	5.36	5.73	.9999
1st fifty.......	10.28	.072	0.59	5.45	.221	.136	4.83	8.00	1.0000
1st hundred...	11.44	.081	0.33	5.70	.228	.095	5.74	7.67	1.0000

Let us take a concrete case. Suppose that we wish to show that the two-point limen upon the forearm is greater than that upon the eyelid. In the Table we find actual determinations of this limen (L) and of h in each of the two cases.[2]

[2] The data are borrowed from the Cornell Laboratory and are for different O's. They are used here merely as an example of a method of treatment for data from a single O. The figures for the forearm were obtained in the quantitative drill-course by two exceptionally careful students, and are based on 100 series of 5 separations each. The figures for the eyelid are unpublished preliminary experiments by E. J. Gates on Observer W (*v.* this Journal, 26 1915, 152 ff.), for whom 100 series of 9 separations each were used. Mr. G. J. Rich has kindly assisted me in the computations. Urban's tables were used.

The computations are made separately for each 10 of the 100 series, and also for the first 20, the first 50, and the entire 100 series. Our problem in each case is to determine the degree of significance which may be attached to the difference between the limens The computation for the first line of the Table is as follows:

For the forearm, $L = 7.55$ mm., $h = .040$. The probable error of the single observation can be found, for $P.\ E. = \frac{0.4769}{h} = 11.92$. The probable error of the limen, $P.\ E._{L}$ (corresponding statistically to the "probable error of the mean"), is $P.\ E.$ divided by the square root of the number of cases. The number of cases in 10 series of 5 separations is 50; but the 50 are not all equally weighted. A weighted value can be obtained by multiplying the sum of the values of P (Urban's tables) by the number of series, 10. In this case, the values of p are 40, 50, 60, 80, and 100%; of P, 0.98, 1.00, 0.98, 0.77, and 0 respectively. The sum of P is 3.73; thus the weighted number of cases is 37.3, and $P.\ E._{L} = \frac{11.92}{\sqrt{37.3}} = 1.96$. In a similar manner $P.\ E._{L}$ for the eyelid is found to be 0.456. The difference between the two limens is $7.55 - 5.70 = 1.85$ mm. The probable error of this difference is the square root of the sum of the squares of the probable errors of the limens, *i. e.*, $P.\ E._{D} = \sqrt{1.96^2 + .456^2} = 2.065$. The ratio of the difference to its probable error is $\frac{1.85}{2.065} = 0.90$. The value of the probability integral for this ratio may be found from a table of the integral to be 0.4539.[3] Thus there are 4,539 chances out of 10,000 that the difference will not deviate from the observed value by an amount greater than the observed value; or $10{,}000 - 4{,}539 = 5{,}461$ that it will so deviate. In only one-half of these 5,461 cases will the deviation be a decrease. Thus the probability that the difference is less than zero (deviates negatively by an amount greater than itself) is 0.2730. From this it follows that the probable correctness of the difference (*i. e.*, the probability that the limen for the forearm is greater than that of the eyelid) is $1.0000 - 0.2730 = 0.7270$.

[3] There are simple tables arranged for this ratio; *v.* Table II in W. W. Johnson, *Theory of Errors and Method of Least Squares*, 1915; and the less complete table in E. L. Thorndike, *Theory of Mental and Social Measurements*, 1913, 200.

Inspection of the column of probable correctnesses in the table shows that in many cases a difference based on only ten series may be highly significant. In the cases of the 6th and 10th series the probable correctness is 'certainty' within the limits of precision of a 4-place table. (In such a table any difference greater than 6 times its probable error is 'certain.') On the other hand, the low probable correctness of the differences for the first and fourth tens would doubtless render these series by themselves unacceptable as a basis for scientific conclusion. It should be kept in mind that 0.5000 is pure chance and is the minimal value obtainable. Thus a probable correctness of 0.7270 has no more significance than, say, a correlation of 0.45.

The last four figures of the last column show that there was relatively little advantage in our continuing these experiments beyond 20 series in order to demonstrate a difference between the two limens. Had the difference been less, more series might have been necessary; had it been greater, 5 or 10 series might have sufficed.

The importance of practice lies not in its effect upon the magnitude of the limen, but in its effect upon constancy of judgment. The reason that the practised limen is a better anthropometric measure than the unpractised limen is not that it is smaller, but that it is less subject to variation.[4] This decrease of variability with practice is evident in Fernberger's h's and in the *M. V.* 's among his observers. Lack of practice gives a small h and consequently a small probable correctness.

Any attempt to lay down rules for work upon the unpractised observer should be based upon a determination of the best number of practice-series as well as the best number of principal series. Fernberger seems to say that it is better to take 50 series than 25 because the interval of uncertainty is much nearer its practised limit in the former case than in the latter (7.90 for 25, 6.89 for 50); but from such an argument one might infer that it would be still better to take the second 25, and discard the first 25 as practice-series, since the interval of uncertainty for the second 25 is less than for the first 50 (5.96 for the second 25). Doubtless Fernberger does not suggest this possibility because he believes that 25 series are too few; but the reasonableness of this contention

[4] Fernberger demonstrates this fact in his paper On the Relation of the Methods of Just Perceptible Differences and Constant Stimuli, *Psychol. Rev. Monog.* No. 61, 1912, 19 ff. Doubtless he means to imply it in the place under discussion.

shows exactly how great is the necessity for always taking the number of cases explicitly into account. It would be interesting in Fernberger's experiment to be able to compare not only the second 25 with the first 50, but also the last 30 and 40 of the 50 with the entire 50, and the last 50 and 75 of the 100 with the entire 100. In this way one might, in any particular case of comparison, determine the total number of observations desirable and the number which should be discarded as practice-series.

Even when practice fails to enter in at all, the problem of the least number of cases reliable remains. It is often desirable repeatedly to measure the same limen for a single practised observer under varying conditions. One may, for instance, wish to know how sensitivity varies with the time of day, with the degree of recovery from a pathological condition, or with casually induced organic or emotional states. Under such conditions the time available for the determination is likely to be limited, and the question then arises: can enough observations be made to give a satisfactory measure? Such a question can be answered only for the particular instance, since the greater the difference in question, the less the necessary number of cases. The greater constancy of the practised observer, however, makes it probable that we could accept fewer cases than would be allowable were we measuring individual differences between unpractised observers.